Adaptive initialization of a EvKNN classification algorithm

Stefen CHAN WAI TIM, Michèle ROMBAUT and Denis PELLERIN

Abstract The establishment of the learning data base is a long and tedious task that must be carried out before starting the classification process. An Evidential KNN (EvKNN) has been developed in order to help the user, which proposes the "best" samples to label according to a strategy. However, at the beginning of this task, the classes are not clearly defined and are represented by a number of labeled samples smaller than the k required samples for EvKNN. In this paper, we propose to take into account the available information on the classes using an adapted evidential model. The algorithm presented in this paper has been tested on the classification of an image collection.

1 Problem positionning

1.1 Classification problem

The classification process needs some a priori knowledge for the class definition. This knowledge can be modeled for the classes (neural network, bayesian classifier) or can be limited to a learning set composed of labeled samples (KNN, SVM). In any case, the classifier needs a learning set to manage the classification of unlabeled samples from the collection and this learning set must be representative of the classes. When it is the case, the classical approaches are very efficient and are used in numerous applications. However, setting up such learning database can be a laborious task for the user.

We proposed in a previous paper [1], an assistance system for image collection classification presented Fig. 1. The first part of the system, based on Evidential KNN (EvKNN), models all available knowledge provided by the already labeled images in order to structure the unlabeled ones. The second part is a user assistance system (based on active learning) that proposes an ordered list of images to be labeled ac-

GIPSA-Lab/DIS, CNRS - UJF, Grenoble FRANCE, e-mail: `firstname.author@gipsa-lab.fr`

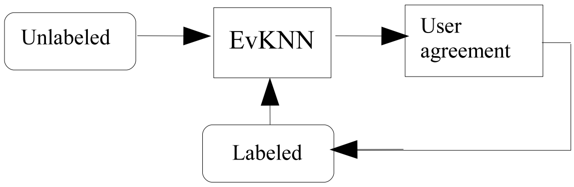

Fig. 1 Labeling process of the training set. At the beginning, the training set is almost empty. The EvKNN classifier takes all available labeled samples to propose to the user a label for an unlabeled sample. With agreement of the user, the new labeled sample is stored in the labeled set.

cording to a specific strategy and assign a possible label. Using a suitable interface, the user agrees or disagrees with the proposal, and the global knowledge is updated.

This paper deals with the beginning of the first part of the labeling process, when the training set is almost empty, with only some labeled samples. In this case, there are generally less than k samples that belong to each known class and the samples are not completely representative of a class. Therefore, EvKNN algorithm cannot be used directly without adaptations. The adaptations are presented in Section 2, and the adapted algorithm is tested on an image collection (Section 3).

1.2 Evidential KNN

In [2], T. Denœux explains that ”voting KNN” procedures show several limitations and he proposes to take into account the distance from the neighbors to model uncertainty and imprecision in class labels. It is assumed that the set of training samples is composed of enough samples for each class of decision. In the KNN algorithm, when there are at least k known samples of each class, there are enough training neighbors to model the membership of every incoming unlabeled sample to each class. T. Denœux proposes to model these memberships by belief functions (see Eq. 1).

We assume that x^s is the incoming unlabeled sample, and x_q^i is a labeled sample belonging to class C_q, one of the Q known classes. $d^{s,i}$ is the distance between these two samples in the parameter space. The knowledge of the x_q^i label gives information about the class of x^s. The basic belief assignment (BBA) $m_i^{\Omega_q}$ is defined on $\Omega_q = \{H_q, \overline{H_q}\}$, where hypothesis H_q means “sample x^s belongs to class C_q”, whereas $\overline{H_q}$ is the opposite hypothesis:

$$\begin{aligned} m_i^{\Omega_q}(H_q) &= \alpha_q . e^{-\left(\frac{d^{s,i}}{\sigma_q}\right)^\beta} \\ m_i^{\Omega_q}(\Omega_q) &= 1 - m_i^{\Omega_q}(H_q) \end{aligned} \tag{1}$$

This model is very interesting when a class is represented by several samples in the parameter space. It means that two distant samples in this space can still belong to the same class. It can be noticed that for a particular class C_q, the proposed BBA form will not cause conflict.

If there are k neighbors x_q^i, we can define k BBAs on the same frame of discernment Ω_q that can be conjunctively combined to give the BBA m^{Ω_q} concerning the sample x^s on membership to class C_q. In our previous paper [1], we proposed some adaptations in the combination. Contrary to Denœux's propositions in [2], we assumed that the Q classes are not exclusive. The combination of the Q BBAs m^{Ω_q} extended to the space $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_Q$ gives one BBA with possible multi-labeling. The combination architecture is described in [1].

1.3 Initialisation step

The EvKNN method is very efficient if the number of classes Q is known, and if the training set is representative enough. If not, the performance of the classifier is reduced. In the later case, the goal is to model the poor information efficiently and possibly to ask an expert to validate the decision. It is also important to take into account the difference of available samples for each class, as well as the relative properties of the classes. The Belief Function Model is particularly well adapted to model such poor information, and given a large mass of belief for sets Ω_q.

In this paper, we describe an adaptive method to propose a decision to an expert. At each step, the choice of the expert is used to improve the knowledge to get a labeled sample and to adapt the information model for the class C_q. At the begining, the training set is only composed of some labeled samples, for instance less than k samples for each known class. The problem is to model this knowledge about the belonging of x^s to a known class. Then, a proposition is made that is validated by the operator.

2 Adaptive model of knowledge

The labeled neighbor x_q^i gives information on the belonging of x_s to the class C_q that can be modeled by the equations 1. The parameter σ_q weights the distance $d_q^{s,i}$ between the sample x_s and the labeled sample x_q^i. The parameter α_q is the discounting parameter that models the unreliability of the source of information. In the classification step, if the distance between two samples is null then it is not completely sure that x_s belongs to the same class C_q of x_q^i. Generally, the two parameters σ_q and α_q are constant, at least for each class. We propose to adapt them using the knowledge from known classes C_q, that is to adapt them according to number and position of labeled samples in the parameter space.

2.1 Adaptation of σ_q

For one unlabeled sample x_s, the k neighbors x_q^i of each class C_q are extracted if they exist. If not, all labeled samples of the class C_q are used. We propose to adapt the distance $d_q^{s,i}$ between x_s and x_q^i by defining a relative distance $\left(\frac{d^{s,i}}{\sigma_q}\right)$. The idea is to take into account the mean distance $d_{q'}^{s,i}$ of x_s from all samples $x_{q'}^i \in C_{q'}$ for all classes $C_{q'}$. We propose to define σ_q where $C_{q'}$ and C_q are known classes and γ is a tuning parameter:

$$\sigma_q = \gamma \cdot min_{q'}(mean_{q' \neq q}(d_{q'}^{s,i})) \tag{2}$$

Therefore in equation 1, the distance $d_q^{s,i}$ is weighted by mean distance to the nearest class $C_{q'}$. The consequence of this definition is :

- if the near class $C_{q'}$ has a mean distance comparable to the distance $d_q^{s,i}$, the doubt is high. This can be modeled with a large mass attributed to each $m_i^{\Omega_q}(\Omega_q)$, given a small value to σ_q.
- if the near class $C_{q'}$ has a mean distance higher than the distance $d_q^{s,i}$, the doubt is low. This can be modeled with a larger mass attributed to $m_i^{\Omega_q}(H_q)$, given a large value to σ_q.

2.2 Adaptation of α_q

The number of known neighbors has a great influence on the BBA's values. If one class C_q contains a lot of labeled samples (more than k), due to the definition of the BBA (Eq. 1), the conjonctive combination of k BBAs reinforce the $m_i^{\Omega_q}(H_q)$. On the contrary, if the class C_q is underepresented ($k_q < k$), then BBA is less informative. This can induce an imbalance between the classes.

We propose to adapt the parameter α_q to the number of known neighbors for each class C_q. The idea is to reinforce the mass $m_i^{\Omega_q}(H_q)$ when $k_q < k$. The definition of α_q is:

$$\alpha_q = \alpha_0^{\frac{1}{1+k-k_q}} \tag{3}$$

where $\alpha_0 = 0.8$. In equation 3, $\alpha_q > \alpha_0$ when the number of neighbors k_q is less than k, to reinforce the mass of the H_q hypothesis. It is equal to α_0 when $k_q = k$.

3 Application to image classification

The automatic classification problem is very complex for image (and video) collections because the user interprets the semantic content. The extracted attributes from the images are not directly connected to the classes wished by the user. During the labeling process of the learning set, the classification system must take into account the knowledge of the user in order to ”learn” the classes C_q. In the KNN approach, the system requires samples of images (or videos) that are labeled by the user. The operation is long and tedious. In a previous work [1], we developed an assistance classification system based on the fact that it is difficult for a user to a priori define all the classes, and manage all the images from the database simultaneously.

3.1 Global architecture of the classification system

It could be difficult for a user to classify a set of images, particularly when the set is large and the classes are not defined a priori. This is the case, for instance, when somebody wants to store his holiday images, not only by time stamp, but also by themes (actions: visit, drive..., locations: at home, outdoor...). The images can be multi-labeled. Rather than submitting all the images simultaneously, or one by one in random order, the idea is to propose an ”adequate” order following a sampling strategy by an active learning process, rarely used for multi-labeling [3]. We retain the main elements of the developed system. The main idea is to select images for the user which are ”interesting” to classify according to a specific strategy and to propose a label. The user can accept the proposed label, or change the label or create a new class. The automatic image selection is carried out from the accumulated knowledge from the previous image classification.

The framework is divided into two main parts [4]: a fully automatic part for “modeling the knowledge” presented in this paper, and another part that concerns the user interactions in order to select the images to be labeled via a graphic user interface. The entire framework is presented as three modules in Fig. 2.

3.2 Sampling strategies

A small set of chosen images is proposed to the user to classify. These could be very similar to labeled images (Most Positive unlabeled images) or very different from labeled images (Most Rejected unlabeled images). We chose the Most Positive strategy for the test because it introduces an imbalance of number of neighbors between classes during the process.

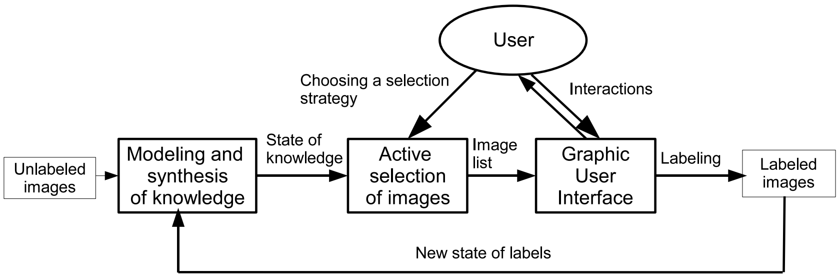

Fig. 2 Architecture of the system of classification

We define a positive hypothesis $\omega_p^q \in \Omega_P \subset \Omega$ composed of only one local positive hypothesis such as H_q, the others corresponding to local negative ones such as $\overline{H_n}$:

$$\omega_p^q = (H_q, \overline{H}_{n_1}, \overline{H}_{n_2}, \cdots, \overline{H}_{n_N}) \tag{4}$$

This positive hypothesis ω_p^q means that the unlabeled image belongs to the single class C_q. The strategy, sometimes named "most relevant" [5], selects the unlabeled images that obtain the highest pignistic probability [6] computed on Ω_P, subset of Ω made up of only positive hypotheses ω_p^q (Eq. 4). It corresponds to the selection of "easy to classify" images, because the visual content is very similar to already labeled images.

3.3 Results

The classification algorithm has been tested on a Corel database of 321 images (Examples in Fig. 3). The database contains 9 classes ('Monuments', 'Bus', 'Dinausors', 'Elephants', 'Mountains', 'Flowers', 'Horses', 'Meals', 'Faces'), and each class has between 15 to 46 images. Some classes are very heterogeneous from the color point of view.

For each image, two kinds of features (color and orientation) have been extracted. For color, classic 3D histograms in HSV domain have been used with 8 bins in each dimension, giving 512 components. For orientation, we used horizontal and vertical gradient filters that give a histogram of 64 bins.

At any time, an unlabeled image is proposed to the user according to the chosen strategy (here the Most Positive) as well as a proposed label. The user can accept the proposed label or reject it. In the later case the proposal is recognized as false proposal. The objective is to limit such false proposals in order to make the task easier for the user. The test is performed automatically since the ground truth is known.

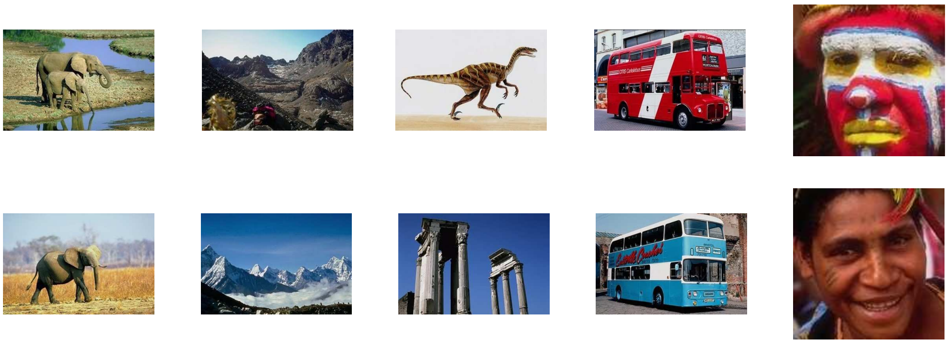

Fig. 3 Examples of color images belonging to the collection

3.3.1 Effect of the parameter σ_q

An example of comparison is given in Fig. 4. We chose $\sigma = 0.5$ (best result) in the constant case, whereas σ_q adapted case follows Eq. 2.

Compared to σ constant, σ_q adapted resulted in a reduced number of false classification proposals. Indeed, if σ_q is too small, the mass goes to the doubt and part of information disappears. If σ_q is too large, the mass $m(H_q)$ tends towards α_q. Here we are too categorical comparatively to the complexity of the content. For σ_q adapted, if the class is far from any other one then σ_q is large, otherwise σ_q is small.

3.3.2 Effect of the parameter α_q

An example of comparison is given in Fig. 5. We chose $\alpha = 0.8$ (best result) in the constant case, whereas α_q adapted follows Eq. 3 with $\alpha_0 = 0.8$.

Compared to α constant, α_q adapted resulted in reduced number of false classification proposals. This result is due to the reduction of imbalance on the masses during the search of neighbors. The value of α_q is close to 1 when the number of neighbors is 1, giving more mass to the H_q hypothesis. It is equal to α_0 when $k_q = k$.

4 Conclusion

The adapted EvKNN proposed in this paper makes the task of the user easier during long and tedious labeling of the training set. The algorithm takes into account the real known neighbors (less than k) and the relative distances of the classes. Because the user is in the loop, a new class can be added when a sample arrives, and in this case, the proposed adapted EvKNN is particularly efficient. The algorithm has been tested on an image collection. The image classification process is very complex

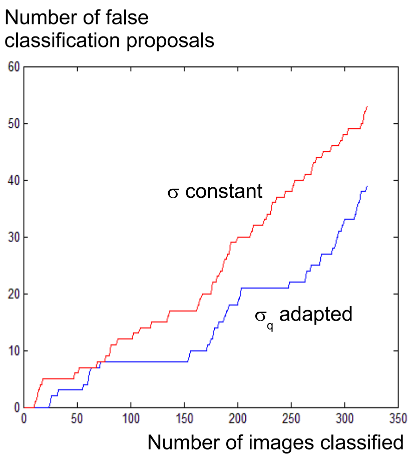

Fig. 4 Comparison of classification for σ constant and σ_q adapted (α constant)

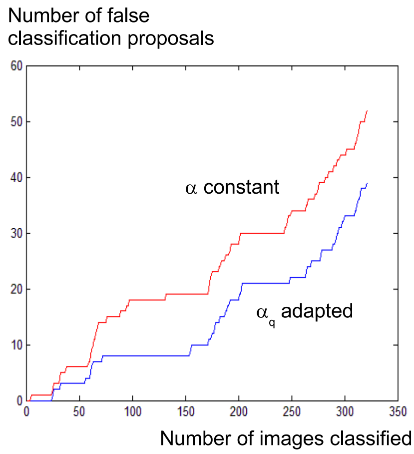

Fig. 5 Comparison of classification results for α constant and α_q adapted (σ constant)

because the user attaches semantic interpretation for an image that an automatic system can not manage using simple image attributes.

Acknowledgements We thank the Rhône-Alpes region for its support with the LIMA project.

References

1. H. Goëau, M. Rombaut, D. Pellerin, O. Buisson, "Multi-labeled image classification by TBM active learning." in *Workshop on the Theory of Belief Function*, 2010,April 1 – 2,Brest.
2. T. Denœux, "A k-nearest neighbor classification rule based on dempster-shafer theory." in *IEEE Trans. on SMC*, 1995, pp. 804–813.
3. G.-J. Qi, X.-S. Hua, Y. Rui, J. Tang, and H.-J. Zhang, "Two-dimensional multi-label active learning with an efficient online adaptation model for image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 99, no. 1, 2008.
4. H. Goëau, "Structuration de collections d'images par apprentissage actif crédibiliste," PhD, Université Joseph Fourier, Grenoble, France, May 2009.
5. M. Crucianu, M. Ferecatu, and N. Boujemaa, "Relevance feedback for image retrieval: a short survey," Report of the DELOS2 European Network of Excellence (FP6), 2004. [Online]. Available: `citeseer.ist.psu.edu/crucianu04relevance.html`
6. P. Smets, "Decision making in tbm: the necessity of the pignistic transformation." *Journal of approximate Reasonning*, vol. 38, no. Issue 2,, pp. 133–147, February 2005.
7. K. Brinker, "Incorporating diversity in active learning with support vector machines," in *Proceedings of the 20th International Conference on Machine Learning*. AAAI Press, 2003, pp. 59–66.